Chapter 8: An Algorithmic Ready-made: Trevor Paglen, *Adversarially Evolved Hallucination* and *Eigenface (Even The Dead Are Not Safe)*

Background: Notions of Transparency

While Blas's work references a notion of opacity, the work of the artist Trevor Paglen conjures up its opposite, transparency. Much of Paglen's work involves the production of images in environments and milieus where there has been a conscious effort to suppress images. Paglen's artwork has exposed the nature of opacity in covert military and political agencies, such as the NSA and the CIA, revealing a surrealist character to their agendas and the physical structures through which they operate, forcing a kind of transparency into their workings. His recent works concern the topic of machine vision and its increasing prevalence in society. He approaches the topic of machine vision as he does his other subject matter, exposing the hidden and obscure processes through which it operates and creating spaces of transparency and visuality. What particularly interests Paglen about machine vision is the growing economy of images, produced by and for machines, that is evolving and, for the most part, going unseen by the human audience. His inquiry acknowledges the influence of machine vision in bringing about a transformation in contemporary visual culture as a whole, where increasingly, he argues, human vision is an exception to the

rule.[1] And yet, as Paglen argues, these machine vision images play an increasingly large role in human interactions, with "their functions changing from representation and mediation, to activations, operations, and enforcement," thus making it necessary for us to scrutinize not only the images but the processes through which they enact a form of vision.[2]

In his attempts to reveal the inner workings of forms of machine vision, Paglen produced a series of images using different artificial intelligence (AI) systems of machine vision. These works culminated in an exhibition at Metro Pictures Gallery called *A Study of Invisible Images*.[3] Some of the works included in this exhibition appropriate the actual processes of algorithms that have been designed to enact different operations of machine vision. Two such works will be discussed here: a series titled *Adversarially Evolved Hallucination* (2017) and a series of works made with the eigenface algorithm titled *Eigenface (Even The Dead Are Not Safe)* (2017).[4] In both of these works, Paglen approaches the technology of image recognition systems (with the latter dealing directly with a facial recognition system) and provides a kind of visual dissection of their algorithmic processes.

In producing these works, Paglen is part of a growing movement of artists using code to make art. I would describe Paglen's artistic approach to image and face recognition technology as internal; that is, he uses the algorithmic processes themselves as material for the

1 Trevor Paglen, "Invisible Images (Your Pictures Are looking At You)," *The New Inquiry*, December 8, 2016, https://thenewinquiry.com/invisible-images-your-pictures-are-looking-at-you/.

2 Ibid.

3 Trevor Paglen, *A Study of Invisible Images*, exhibition, September 8-October 21, 2017, Metro Pictures Gallery, New York, NY.

4 Paglen included other works in this exhibition that refer directly to different facial recognition systems, including *It Began as a Military Exhibit* and *Machine-Readable Hito*. I chose the two works discussed here because of their connection and relevance to the overall analysis. Another of Paglen's works that bears on facial recognition (specifically automated facial-behavioral analysis) is his piece *Sight Machine*, performance, January 14, 2017, San Francisco, CA.

production of images. In this way, his artwork is less a confrontation with the technology's front-end operations and more an engagement with the internal workings of how the technology enacts recognition. This approach in turn, alters the production and operation of the algorithm itself. This contrasts with Blas's work, where our attention is largely drawn toward a confrontation between the recognition device and the embodied subject under surveillance. In presenting images that express how the internal workings of AI systems see, Paglen's work has a way of being pedagogical, teaching an observer of his images about how an algorithm sees. In the absence of further explanation, this aspect can make his work somewhat opaque and hard to penetrate.

One of the strengths of Paglen's work is its ability to put technological processes into dialogue with discourses in art history and political philosophy. In doing so, his works not only expose but also actively reimagine the technological processes at issue and hold out the promise of a reinterpretation of the possibilities of the technology's operations. Paglen finds unintentional affinities between abstract art and linear classifiers – the archetypes used to train algorithms how to recognize objects. I see Paglen's use of algorithmic processes to produce images within a cultural context, in fact, as a reference to another art-historical practice, that of the ready-made, a method of art production involving found but modified objects/products, first used by Marcel Duchamp, who presented manufactured objects as objects of art. An aspect of the ready-made is the use of ubiquitous, everyday objects. In the contemporary world, algorithms have become one such everyday object; although they are immaterial and their presence undetectable, they are ubiquitous at all levels and areas of interaction. Paglen's appropriation of image recognition algorithms treats them as an immaterial ready-made: that is, he uses preprogrammed or "found" algorithms and modifies their output, positioning this output within a discourse of art production and within various conceptual frameworks. In producing an "algorithmic ready-made," Paglen takes the everyday object of the algorithm and merges its technical processes with the conceptual.

Adversarially Evolved Hallucination (2017)

Paglen produced the series of works titled *Adversarially Evolved Hallucination* during a residency at Stanford University. In referring to these images as "hallucinations" generated from machine-vision processes, Paglen characterizes them as departing from reality – as though we are about to see what machines see while on acid. There is, he is saying, an illusory quality to the machine-vision processes usually deemed objective, scientific and engaged in definitive operations. The word "adversarial" in the title of the series refers specifically to images that are developed by computer scientists to trick image recognition systems. These "adversarial images" are developed to exploit weaknesses in the recognition algorithm, making them see things that are not there. The purpose of these images is to better train the algorithm – a kind of exercise of productive technical failure. These adversarial representations eventually get incorporated into the training sets used by the algorithm in machine learning. With the help of a small team of developers, Paglen customized the software used for recognition and developed software suites. He used two separate algorithmic processes: one that taught the machine to recognize objects through training sets and the other that was used to generate adversarial images based on the same training set. Paglen referred to the secondary algorithm as the "Generator,"[5] while the first he called the "Discriminator."[6] Paglen designed these two algorithms to engage in a dialogue with each other: to play a game in which the second algorithm attempted to fool the first through the production of adversarial images. Paglen explains:

> The two AIs go back and forth thousands or millions of times, until the Generator has learned how to make images that can reliably fool the Discriminator. The images that come out of this process are called Hallucinations. Together, the AIs have evolved an image that

5 Metro Pictures, *Trevor Paglen: A Study of Invisible Images*, (New York: Metro Pictures, 2017), 24.

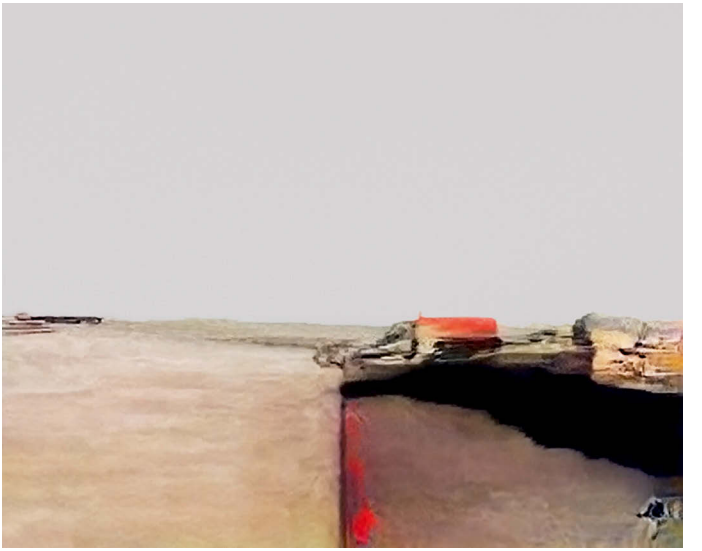
6 Ibid.

> is entirely synthetic and has no referent in reality, but that the pair of AIs believe are examples of things they've been trained to see.[7]

The resulting images are surreal, pixelated forms that emerge in vivid color. The game that the two AIs play is not only a way of exploiting the weaknesses and parameters of algorithmic learning; it is also a game that plays with the idea of recognition itself. How does the AI know something as it is? What can be recognizable? In this algorithmic dialogue, Paglen toys with the relationship between visual form and knowledge. A key aspect of these works is found in their titles, which are taken from the categories of training set images used by the algorithms. Instead of the phrase "training set," Paglen uses the term "corpus," bringing attention to the role of training sets as a body of knowledge. The corpuses that are referred to in the titles include "Omens and Portents," "Monsters of Capitalism," "American Predators" and "The Aftermath of the First Smart War." These categories organize the types of objects being recognized within Paglen's software suite, and they clearly differ from the actual categories of training sets used in machine image recognition. Instead, they reflect concepts that act as metaphors, describing the wider socio-political contexts in which the technology is used. The images that result from this algorithmic game of recognition render these metaphors visible. For example, one image, *Highway of Death (Corpus: The Aftermath of the First Smart War)* (figure 27), references the militaristic context of the use of image recognition. The image is a surrealistic rendering of a desert battlefield with no humans present. What is present in the image is a shift in replacing the categories of training sets to recognize objects or people to the recognition of the cultural contexts of its operation. Through Paglen's production of adversarial images, he is experimenting with training an algorithm (and the human observer of the generated image) to not only see objects but also to see concepts that shed light on the wider cultural contexts in which these technologies intervene.

7 Ibid.

Figure 27: Trevor Paglen, "Highway of Death (Corpus: The Aftermath of the First Smart War)" Adversarially Evolved Hallucination, 2017. dye sublimation metal print. Image courtesy of the artist



For Paglen, the importance of developing a vocabulary through metaphor connects back, I think, to a previous series of his, *Symbology* (2006), in which he collected material paraphernalia relating to black ops and covert activities of the CIA, including fabric patches, designed by members of covert units, that are embroidered with coded insignia, symbols such as dragons, arrows, animals, planets and phrases such as "We Own the Night." The character of the metaphoric forms and titles found in the images of *Adversarial Hallucinations* may be seen as mirroring these ominous symbols and phrases. Both reflect a world that would otherwise not be seen, and once exhibited become objects of conceptual inquiry. Paglen describes the patches used in *Symbology* as follows: "A part of the military's everyday culture […] If you could begin to learn its grammar, you could get a glimpse into the secret world itself."[8] This description

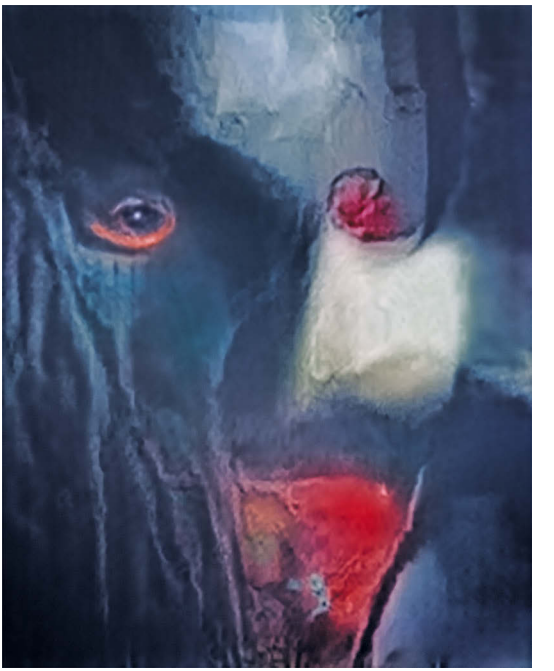
8 Trevor Paglen, *I Could Tell You But Then You Would Have to Be Destroyed By Me: Emblems From the Pentagon's Black World* (Brooklyn: Melville House, 2010), 4, 5.

can be extended, I think, to his approach in *Adversarial Hallucinations*, which develops a visual vocabulary of algorithmic processes, building on a taxonomy of corpuses, providing a grammar to see how machines see within the contexts in which they operate.

Some of the titles of these corpuses refer to literary metaphors. Through the corpus titles, Paglen refers to a conceptual taxonomy based in literature and philosophy. As I have mentioned, one of the corpuses references the allegorical monsters of capital,[9] that is, monsters that have acted as literary metaphors for capitalism. One in particular, takes the form of a face – Paglen's piece *Vampire (Corpus: Monsters of Capitalism)* (figure 28). The image is pretty terrifying. It is of a face that appears to be constructed from melting pixels. It contains all the facial landmarks: eyes, nose, cheeks and mouth. Yet these landmarks are composited together from disparate sources and fragments of light and shadow, creating a kind of digital collage work. The left eye resembles a photograph of an actual eye seemingly cut out from a magazine, and it peers out, furtively, from a dark curtain of velvet blur. From this eye, bluish veins appear to drip down. The other eye, in contrast, is constituted by a dark swirl resembling a wilting red carnation. A jagged lightning bolt parts light and shadow across the face and acts to form the nose. In place of a mouth is a kind of striated, conglomerate of blood-red, pixelated ooze. Much of the face is hidden in shadowless depths of darkness. In the areas of light, specifically in the forehead, one can detect a painterly effect of the multiple layers of digital imagery that constitute this composited face. The predominance of a Dada-like palette of black, white and red contributes to the stark nature of this vampire face. The more realistic left eye is the only part of the face that is a recognizable, complete form. Without that eye one could hardly see a face in this image. That eye is thus both an organizing feature and also the source of the image's terror. For, in its fully realized form, it is what allows the vampire/monster in the image to look back at the observer, and as an observer we feel very much under its mono-gaze.

9 For more on this subject see David McNally, *Monsters of the Market: Zombies, Vampires, Global Capitalism* (Chicago, IL: Haymarket Books, 2012).

Figure 28: Trevor Paglen, "Vampire (Corpus: Monsters of Capitalism)" Adversarially Evolved Hallucination, 2017. dye sublimation metal print. Image courtesy of the artist



Paglen has described how he thinks of AI itself as a monster of capitalism. The subject of the image in *Vampire (Corpus, Monsters of Capitalism)* alludes to Karl Marx's vampire of dead labor: "Capital is dead labour, that, vampire-like, only lives by sucking living labour, and lives the more, the more labour it sucks."[10] As such, the image references a wider context in which market forces drive image-recognition algorithms, and AI in general, to be deployed in order to replace human labor. In his allusion to Marx and with his depiction of a vampire as one particular monster of capitalism, his work may be seen as a critique of the forces and agendas behind the technology's development and use – politically, commercially and militarily.

10 Karl Marx, *Capital: A Critique of Political Economy*, vol. 1, trans. Samuel Moore and Edward Aveling, (Moscow: Progress Publishers, 1887), 163.

His piece visualizes the "face" of this "monster of capitalism," and as such it stands out in the context of this thesis on facial recognition. For this is another construction of a portrait by an algorithm, yet it is one based on and born entirely from a concept. In this way, Paglen also says something about faces as conceptual constructs. This face was constructed, as were the other images in the series, from a training set of thousands, if not millions, of collected images – in this case, various images of vampires.

Figure 29: Trevor Paglen, training set images for "Vampire (Corpus: Monsters of Capitalism)" Adversarially Evolved Hallucination, 2017. Image courtesy of the artist

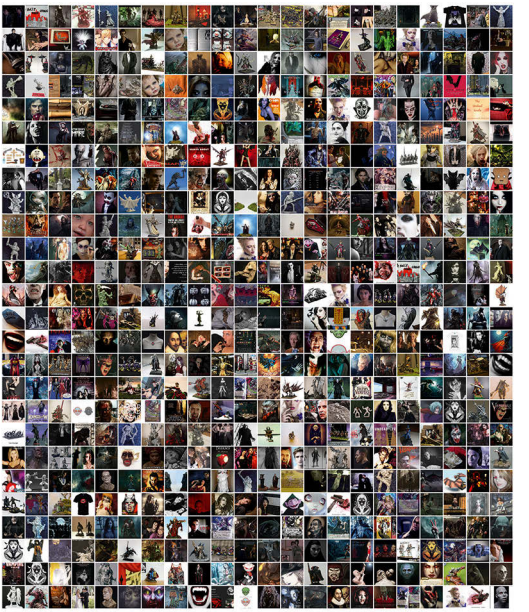


Figure 29 shows a sample of the training library, revealing a wide variety of cultural imaginations of vampires, ranging from the beloved Count von Count Muppet from Sesame Street, to photographs of infant vampires, to depictions of vampiric motifs in fashion spreads and medieval paintings.

In focusing on the initial training-set phase of the recognition process, Paglen problematizes a central issue with automated recognition through AI: the way algorithms learn through training sets and, as such, fix meaning in the physical world. Paglen describes the training-set process as follows:

> [t]his all takes place largely for the most part imperceptibly. Assignations of meanings, of gender, race, species, marketability, and criminality, are done both autonomously and invisibly; we cannot see how we are being named, much less audit the processes through which that naming happens.[11]

In short, as Paglen states, "he who controls the training sets controls the meanings of images."[12] The naming of the categories of training sets ultimately determines the kinds of knowledge an algorithm can (and cannot) produce. Playing with the naming of the corpuses not only allows Paglen to introduce metaphors that describe the contexts of the technology's use but also exposes a loosening in the interplay of signs, both visual and linguistic, that is occurring through the technology. The algorithmic game that produces the images in *Adversarially Evolved Hallucination* bears on notions of recognition, creating a space of negotiation between what is known and what can be seen/recognized. Through the use of language and metaphor, Paglen presents not only the socio-political contexts of the technology but also to suggest that human culture and imagination can intervene in and counter the automated naming processes that ascribe meaning, processes that otherwise remain hidden within the internal, circulatory processes of the technology. Through language and visual metaphor, Paglen injects cultural meaning and context into machine-vision processes. In doing this, he also highlights a central issue of A.I.: that cultural meaning and context are two major blind spots of machine-vision processes.

---

11 Trevor Paglen, "Machine Realism," in *I Was Raised on the Internet*, ed. Omar Kholeif (Munich: Prestel, 2018), 118.

12 Ibid, 116.

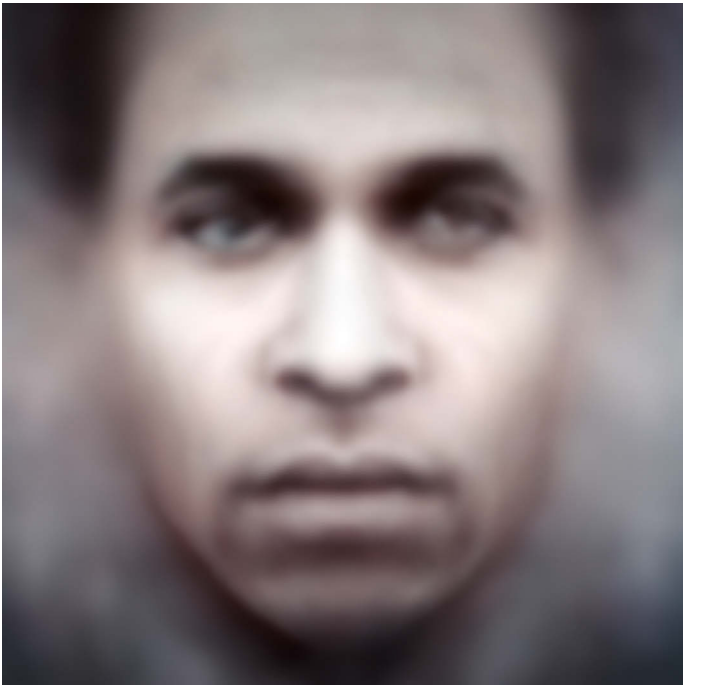
Eigenface (Even the Dead Are Not Safe) (2017)

Paglen elevates the eigenface image to the level of an artwork in his series *Eigenface (Even the Dead Are Not Safe)*. His eigenface portraits are not formed from an average of multiple faces but rather from a compilation of multiple facial images of the same subject. For his subjects, he chose philosophical and literary figures such as Franz Fanon, Samuel Beckett and Simone Weil. (He also made one of the actress, Winona Ryder.) His process of making these eigenface images involved projecting eigenvectors made of each individual subject on to face spaces of other people and mathematically calculating the differences between them. In this way, Paglen reenacted the usual recognition process of the eigenface method, but instead of projecting a captured image of an unknown face, he projected an eigenvector of a known subject, and the resulting differences, which are normally translated into code and stored in a database as representing the identity of the subject, he instead visualized in these images. Because of the way it depicts the features that distinguish the subject from everyone else, Paglen refers to the resulting image as a "faceprint."[13]

As an artwork, Paglen's eigenface portraits contrast with the eigenface images that normally result from the operation of its algorithm, examples of which I have discussed in previous chapters. In the first instance, these portraits are in color. Rather than the usual gray-scale images, these portraits have a washed-out, cold, sepia tone. As they are compilations of multiple images of a single subject, these portraits also clearly portray that subject, albeit with the blur characteristic of eigenface images. The aesthetic of the mathematically abstracted blur, together with the washed-out tones, produces a ghostly rendering of these subjects. The image of Franz Fanon, titled *Fanon (Even the Dead Are Not Safe)* (figure 30), is, I find, the most striking of these portraits. It is a haunting portrayal of the ghost of Fanon, with his face veiled in a chalky pallor and the color of his skin only retained around the shadows of his eyes and lips and along the edges of his face.

13 Metro Pictures, Trevor Paglen: A Study of Invisible Images, 9.

Figure 30: Trevor Paglen "Fanon" Eigenface (Even the Dead Are Not Safe), 2017. dye sublimation metal print. Image courtesy of the artist



There is an expression of urgency in his return gaze. Upon staring into his eyes, one is unsettled by the fact that the color differs slightly in each – one with a bluish hue and the other brown. Fanon once stated: "We who come from the Antilles know one thing only too well: Blue eyes, the people say, frighten the Negro."[14] This portrait thus appears as something of a warning from the past, like Walter Benjamin's take on Paul Klee's *Angelus Novus* as an angel of history; the ghostly gaze of Fanon looks back at us from a rendering of his likeness created by systems of surveillance, as something to fear.

The warning present in the Fanon image is also heralded by the parenthetical remark in its title (which appears in the titles of all

14 Franz Fanon, *Black Skin, White Masks*, trans. Charles Lam Markmann (London: Pluto Press, 1986), 29.

the images in the series): *Even the Dead Are Not Safe*. Given that he is rendering images of historical figures, Paglen's use of this phrase seems apt. Yet this phrase also conjures up a passage from Benjamin's last major piece of writing, "Theses on the Philosophy of History," a text in which he articulates a critique of historicism and the method of Marxist historical materialism. To put it briefly here, Benjamin critiques the idea of history as a continuous path toward progress, and he instead articulates an understanding of history as a memory that arises in a moment of danger. In the same passage, he alleges historicism has been used for the benefit of the "the ruling classes." He states: "In every era the attempt must be made anew to wrest tradition away from a conformism that is about to overpower it […] Only that historian will have the gift of fanning the spark of hope in the past who is firmly convinced that *even the dead* will not be safe from the enemy if he wins."[15] In this passage, Benjamin describes the ways in which the past may be articulated in the present and have a way of threatening both. The representation of the figures in Paglen's eigenface images can be understood to invoke Benjamin's warning.

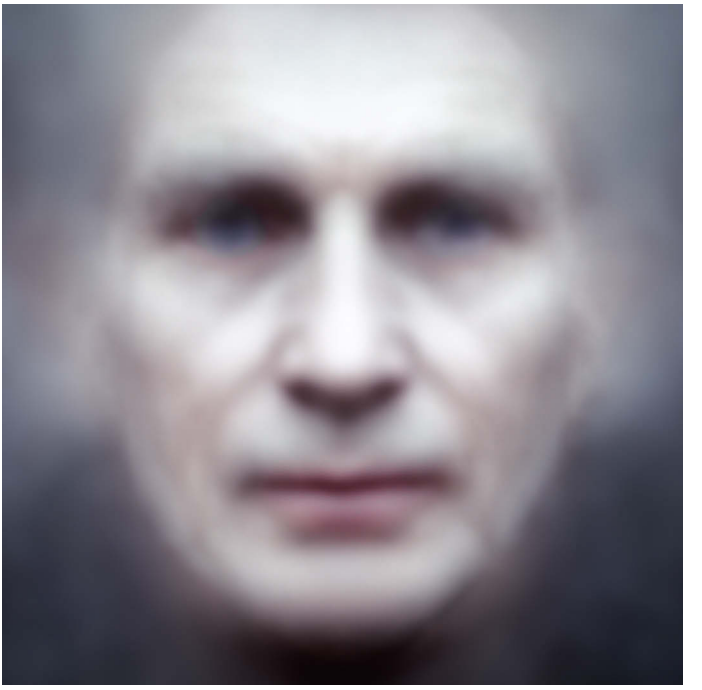
The work of Fanon has been referenced in connection to critiques of biometric systems before. In surveillance scholar Simone Browne's study of contemporary surveillance practices and the historical archive of transatlantic slavery, she borrows from Franz Fanon's term epidermalization in order to coin the term "digital epidermalization," which she defines as "the exercise of power cast by the disembodied gaze of certain surveillance technologies […] that can be employed to do the work of alienating the subject by producing a truth about the racial body and one's identity (or identities) despite the subjects' claims."[16] Browne's use of the term defines a denied subjectivity that is produced through the application of biometric recognition. Paglen's eigenface of Fanon visualizes this denied subjectivity with a representation of Fanon produced through the very systems of this disembodied gaze. As Paglen brings forth Fanon's figure in this piece as a ghostly warning, an invitation of di-

15 Benjamin, "Theses on the Philosophy of History," 255 (italics in original).

16 Browne, *Dark Matters*, 110.

alogue between Fanon's discourse and the dialectics of recognition emerges.

Figure 31: Trevor Paglen, "Beckett" Eigenface (Even the Dead Are Not Safe), 2017. dye sublimation metal print. Image courtesy of the artist



The portraits of Beckett and Weil (figure 31 and figure 32), like the Fanon portrait, present their subjects with a neutral expression and in a forward-facing pose. Fixed as an eigenface image, the head of each figure floats ominously in a sea of pixelated blur. What to make of these ghostly heads of dead revolutionaries and philosophers? As with the production of his images in the *Adversarial Hallucinations* series, with his eigenface portraits Paglen customizes the training set category; he chooses his subjects, all of whom stand out as archetypes of critical thinkers. Weil's face, however, wears the expression of a slight, Mona Lisa-like smile. This is perhaps because, as a woman, she has been expected to smile and to generally present

a pleasant facial expression that is not expected of men. These are, after all, statistical averages of a collection of facial images of each figure, and as such they capture something of the public expression of each subject.

Figure 32: Trevor Paglen, "Weil" Eigenface (Even the Dead Are Not Safe), 2017. dye sublimation metal print. Image courtesy of the artist

In their eigenface images, these three figures are transformed into archetypes in two ways. They are each archetypes of a particular kind of critical thinker. The work of each offers critical analyses of systems of power that speak to the contexts of facial recognition and its processes of meaning production. Beckett, for example, in his work as a novelist, playwright and poet, represents an attack on the realist tradition; he dispenses with the narrative element of a unity of time and place, instead focusing on revealing the essence of the human condition. Weil's work as a philosopher and as a polit-

ical activist fighting for the interests of the working class included a critique of forms of oppression, whether the exploitation of capitalism or the elitist bureaucracy of Marxism. The eigenface images of each present a visual rendering of each figure as an archetype of critical thought. Yet, through the gender and race of each figure, these portraits also present archetypes of institutional categories of identification, an association I think is intentional on the artist's part. As such, we can surmise that "even the dead are not safe," that is, not safe from their inclusion within a categorical "type" as part of the identification practices of AFR technology.

In producing portraits of subjects whose work precedes their image, and whose work also contributes to a critique of the notions of power and identity, Paglen inverts a truth of recognition through automated processes. In bringing these figures "back from the dead" through the form of an eigenface classifier, he brings together the discourses of critical thought in philosophy and contemporary systems of institutional enforcement. The critiques of power that are embodied in these figures are tools. In the course of a discussion of algorithmic prediction in facial recognition, Paglen mentions: "An analysis of power, of capital, of race, [these] are the kinds of tools that you bring to the conversation that are ultimately more helpful."[17] Of the eigenface portraits, he states: "someone like a Fanon or a Weil contributed to social progress precisely by breaking the law – because they were unjust laws […] It's asking whether the development of these technologies will preclude people like Simone Weil or Franz Fanon from ever existing again."[18] Paglen's question is whether, in a society that is increasingly organized through metadata and predictive systems and, as such, enforces certain kinds of "normative behavior," any revolutionary thought and action can survive. Paglen's engagement with this problem involves reference to historical figures and thereby speaks to a precedent of the technology. In other words, there is an element of answering the future

17 Charlie Robin Jones, "The Artist Trevor Paglen, The Surveillance State, and Your Life," SSENSE, May 15, 2019, https://www.ssense.com/en-us/editorial/art/this-artist-the-military-industrial-complex-and-your-life.

18 Ibid.

(both the predictive aspect of the algorithm and a near-future society of increasing automation) with the past. Paglen intentionally exploits the "ghost-face"[19] character of eigenface images as an aesthetic in these portraits, as a way of communicating how the subjects and their legacies of thought haunt the present. Paglen conjures up these figures not only as ominous reminders of the unequal power structures that underlie facial recognition technology but also as a reference to the present necessity of tools of critical discourse.

Concluding Remarks

Through both *Adversarially Evolved Hallucination* (2017) and *Eigenface (Even the Dead Are Not Safe)* (2017), Paglen gets inside some of the issues at the heart of the design of automated recognition technology. Where other scholars and artists have scrutinized the normative categories that guide facial recognition technologies, Paglen challenges how these normative categories are constructed through the technology. In appropriating the training-set phase of the image recognition algorithm, Paglen arrives at one of the ways in which the technology ascribes meaning in the world. Through his imagery he creates a space for dialogue between the dialectics of recognition and the modes of discourse in critical thinking. Together, these approaches situate the technological developments of machine vision technologies within a framework of cultural and political thought. In doing this, he also articulates an existing gap between the design and development of machine-vision systems and the cultural and political worlds in which they are deployed. His images suggest to the observer that the implications of the technology are open ended and as images, they are open to interpretation. As such, he highlights an inherent fluidity of human perception in the face of the machinic gaze.

19 Çarikçi and Özen, "A Face Recognition System," 122.